Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v18n1p165

Determinants of Income Inequality: A Compositional Data Approach

By Hijazi, Zoubeidi

15 March 2025

Electronic Journal of Applied Statistical Analysis
Vol. 18, Issue 01, march 2025, 165-182
DOI: 10.1285/i20705948v18n1p165

Determinants of Income Inequality: A Compositional Data Approach

Rafiq Hijazi*[a] and Taoufik Zoubeidi[b]

[a]*Department of Mathematics and Statistics, Zayed University, Abu Dhabi, UAE*
[b]*Faculty of Health Sciences, University of Ottawa, Ottawa, ON, Canada*

15 March 2025

The determinants and correlates of income distribution have received significant attention in economics and public policy literature over recent decades. Income distribution, representing the share of income received by each quintile or decile of a population expressed as a vector of nonnegative proportions that sum to one, is inherently compositional data. However, most research has traditionally used aggregate inequality measures, such as the Gini coefficient, as the dependent variable when modeling relationships with economic indicators. Unlike a compositional data analysis (CoDA) approach, this reliance on aggregate measures limits insights into the tradeoffs among income classes as inequality determinants change. To date, only one study has applied a logratio-based model to analyze the determinants of income inequality in the U.S., leaving substantial gaps in understanding the broader implications of CoDA in income studies. To address this, our study proposes a Dirichlet regression model for country-level income distribution, integrating relevant economic and development indicators. This model aims to identify key determinants of income inequality and assess their specific impacts on income shares across different income groups. The performance of our proposed model is compared against a traditional Gini-based model, highlighting its potential for more nuanced and comprehensive insights into income distribution dynamics.

keywords: Compositional data, Income distribution, Income Inequality, Dirichlet Regression.

*Corresponding author: rafiq.hijazi@zu.ac.ae

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

In economics and public policy studies, there is an extensive literature on the determinants and correlates of the income distribution and inequality across and within countries (Polacko, 2021; Halkos and Aslanidis, 2023). Income distribution refers to how the total income in a society or an economy is divided among its population and it is presented as the share of income received by each quintile (or decile) of the population (World Bank, 2022a). These income shares are non-negative values that sum to 100% which qualifies them as compositional data. By definition, a composition is a vector of nonnegative proportions representing parts of a whole, subject to a unit-sum constraint.

Most studies modeling the impact of socioeconomic indicators on income distribution use aggregate measures of income inequality - such as the Gini coefficient - as the dependent variable. While this approach can identify significant factors affecting the overall level of inequality, it does not provide insight into the tradeoffs among various income classes in response to changes in the determinants of inequality. By contrast, a compositional data approach, which utilizes the vector of income shares as the dependent variable, enables the study of these tradeoffs directly. Despite the advantages of this approach, our literature review found limited application of compositional data analysis in this field. To date, only one study (Kagalwala et al., 2021) has employed a compositional data analysis (CoDA) approach to explore the relationship between income distribution and its correlates, specifically applying the logratio model of Aitchison (1986) to analyze inequality determinants in the US.

In this research, we propose a Dirichlet regression model to examine country-level income distribution in relation to socioeconomic factors. This approach offers a more nuanced understanding of the determinants of income inequality across countries, capturing shifts within the income distribution itself. The present paper is organized as follows: the literature review on income inequality and compositional data analysis is presented in Section 2. Section 3 presents the models and data description while the empirical results and discussion are given in Section 4. Finally, the concluding remarks are presented in the last section.

2 Literature Review

2.1 Income Inequality

2.1.1 Income Inequality Measures

Income inequality is a multifaceted phenomenon measured by various indicators that address different aspects of income distribution. These include traditional measures like the Gini coefficient, Lorenz curve, and income shares, as well as more sophisticated indices such as the Theil and Atkinson indices. Recent advancements, such as the Zenga-type relative measure, aim to overcome the limitations of classical approaches by focusing on distributional nuances.

The Lorenz curve, developed by Max O. Lorenz in 1905, is the most commonly used

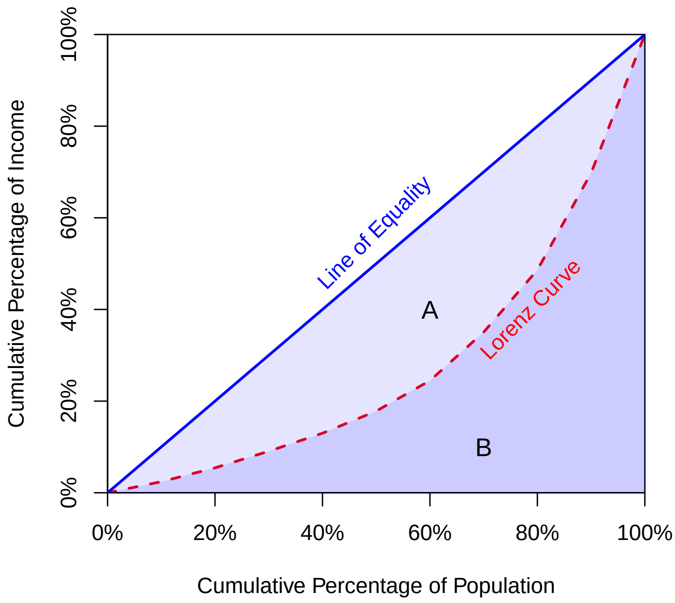

Figure 1: Lorenz Curve

graphical representation of income distribution which is a plot of the cumulative percentage of income against the cumulative percentage of the population as shown in Figure 1. The diagonal line represents the ideal and perfect equality while a curve coinciding with the horizontal axis indicates a perfect inequality. In reality, the Lorenz curves of different countries lie between these two extreme scenarios (Ata et al., 2019). The further the Lorenz curve is away from the line of perfect equality, the more unequal the distribution of income.

In conjunction with the Lorenz curve, different measures have been used to quantify the income inequality. The most widely used is the Gini coefficient, introduced by Corrado Gini in 1912, which measures the extent to which the distribution of income deviates from the ideal income distribution (Ceriani and Verme, 2012). Referring to Figure 1, the Gini coefficient (G) is computed as the ratio of the area between the Lorenz curve and the line of perfect equality (A) to the total area beneath the line of perfect equality ($A + B$), i.e.,

$$G = \frac{A}{A + B} \times 100.$$

A higher Gini coefficient represents a larger income inequality where 0 corresponds to perfect equality and 100 to perfect inequality. According to the World Bank (2022a), between 2008 and 2021, the Gini coefficient ranged between 23.2 (Slovak Republic in 2017) and 63.4 (South Africa in 2017). The Gini coefficient allows for direct comparisons of income distributions. However, as noted by Afonso et al. (2015), very different income distributions could have the same Gini coefficient, which represents a major limitation of this coefficient. Additionally, the Gini coefficient has been criticized for several reasons, including limitations in its measuring capacity (Liu et al., 2020; Osberg, 2017; Gastwirth, 2017).

Beyond the Gini coefficient, several other, but less common, income inequality measures were used in the literature such as the Decile ratio, Robin Hood Index, Generalized Entropy Index (Theil's T and Theil's L), Theil MLD index, Atkinson Index, Palma ratio, etc. For a discussion of these measures, their characteristics, strengths and weaknesses see, for example, De Maio (2007), Haughton and Khandker (2009), Afonso et al. (2015), Liao (2022) and Cobham and Sumner (2013).

In addition to traditional measures of income inequality, the Zenga measure of income inequality is an index that focuses on the distribution of income, particularly highlighting disparities between different population segments, especially at the lower and upper ends (Zenga, 2007). This index offers a more nuanced comparison between the poorest and richest segments, revealing inequalities that the Gini index may overlook (Greselin et al., 2021). The Zenga measure has been widely applied in studies of income distribution across various socio-economic groups, including in Poland and Italy (Jedrzejczak, 2012; Porro and Zenga, 2020). Recent modifications to the Zenga index have been explored by Greselin and Zitikis (2018) and Davydov and Greselin (2020), further enhancing its application and interpretability in assessing income inequality.

The diversity of measures reflects the multidimensionality of income inequality. While traditional measures like the Gini coefficient and Lorenz curve remain widely used, newer indices like the Zenga-type relative measure provide deeper insights into specific aspects of distribution. Combining traditional and advanced measures allows for a more nuanced understanding of inequality and its implications for policy.

2.1.2 Determinants of Income Inequality

The determinants and correlates of income inequality have drawn significant global attention over the past two decades, especially as globalization has brought rising concerns about growing income disparities. Identifying the factors that drive income inequality is essential for understanding the forces shaping national income distribution and for guiding policymakers in developing strategies to enhance citizens' well-being. Ata et al. (2019) indicated that income composition varies widely across regions, countries, and areas due to their unique socioeconomic context. Analyzing these diverse determinants is thus crucial for understanding disparities within and across countries. These factors encompass various categories, including demographic, macroeconomic, financial, and institutional influences (see, e.g., Furceri and Ostry (2019); Kunawotor et al. (2020); Shao (2021); Sidek (2021)).

The growing focus on the drivers of income inequality is reflected in an expanding body of research, examining both developing (see, e.g., Ata et al. (2019); Carvajal et al. (2019); Kunawotor et al. (2020); Asogwa et al. (2022); Nantob et al. (2015)) and developed countries (see, e.g., Roser and Cuaresma (2016); Ngamaba et al. (2018); Tridico (2018); Atkinson (2015); Hoffmann et al. (2020)). In addition, numerous studies have explored specific economies, such as the United States (Kollmeyer, 2018; Kagalwala et al., 2021), United Kingdom (Dorling, 2015), China (Skare et al., 2021), Brazil (Signor et al., 2019), Turkey (Filiztekin et al., 2015), Australia (Gaston and Rajaguru, 2009), and India (Ganaie et al., 2018). Recent comprehensive reviews by Furceri and Ostry (2019) and Shao (2021) provide an overview of the key determinants of income inequality across and within countries. Generally, income inequality reflects disparities in the distribution of wealth and resources within populations, shaped by a range of health, education, economic, and institutional factors. These determinants interact in complex ways, influencing individual opportunities and societal outcomes.

Health disparities, captured by indicators such the *adult mortality rate*, significantly contribute to income inequality. The impact of adult mortality rates on income inequality is not straightforward and varies significantly across different contexts and time periods. While some studies show a positive correlation (Ross et al., 2005), others find a negative or insignificant relationship (Rebeira et al., 2017; Hu et al., 2015). High mortality rates, typically linked to limited healthcare access and inadequate public health infrastructure, disproportionately affect low-income populations by reducing productivity and hindering economic mobility. Conversely, countries with strong healthcare systems and lower mortality rates tend to exhibit narrower income gaps, as healthier populations can more effectively contribute to economic growth.

Beyond health, labor market dynamics significantly influence income inequality, with *wage and salaried workers* playing a pivotal role. Studies generally agree that a higher proportion of wage and salaried workers is associated with reduced income inequality, though the effect size tends to be small (Szymańska and Zielenkiewicz, 2022; Erauskin, 2020). However, this relationship varies across regions; for example, in new EU member states, an increase in labor share has been paradoxically linked to a rise in the Gini coefficient (Šoltés et al., 2023). Conversely, high *unemployment rates* exacerbate income inequality by restricting income sources and increasing dependence on often inadequate social safety nets (Cysne and Turchick, 2012; Petrakos et al., 2023). These patterns highlight the complex and context-specific interplay between labor market structures and income disparities.

Moreover, education emerges as a key driver in reducing income inequality, with higher *primary school enrollment* significantly narrowing income disparities. Primary education equips individuals with foundational skills, enhancing productivity and earning potential, thereby addressing socioeconomic inequality (Shahabadi et al., 2018). This impact is amplified by *government spending on education*, particularly in less developed countries and regions with higher initial inequality levels. Investing in secondary and tertiary education fosters upward mobility and promotes economic equity, resulting in substantial reductions in inequality (Celikay et al., 2016; Płatkowski and Lechman, 2024).

In addition to the aforementioned drivers, economic variables such as per capita GDP, GDP growth, inflation, and government consumption expenditure significantly influence income inequality. Higher *per capita GDP* is generally associated with reduced inequality, although this relationship can be complex and context-dependent (Gil-Alana et al., 2019; Nguyen et al., 2024). Economic growth also has a nuanced impact on income inequality. While some studies find that higher *GDP growth* reduces inequality (Choi, 2006; Gil-Alana et al., 2019), this effect depends on how the benefits of growth are distributed across the population (El Aboudi et al., 2024). *Inflation*, on the other hand, disproportionately affects lower-income households, eroding purchasing power and widening income gaps (Marrero and Rodríguez, 2016). Meanwhile, equitable *government consumption expenditure* can mitigate inequality, while investment effectively boosts growth and curbs inflation but may exacerbate long-term income inequality (Anderson et al., 2017; Sidek, 2021).

Finally, governance quality, particularly *control of corruption*, plays a critical role in reducing income inequality. Corruption entrenches inequality by concentrating wealth among elites and undermining social services. Conversely, effective governance, transparency, and the rule of law are essential for achieving a more equitable income distribution (Adams and Klobodu, 2016; Kunawotor et al., 2020).

2.1.3 Modeling Income Inequality

The majority of studies on income inequality have used the Gini coefficient as the dependent variable, while a smaller number have employed the less common indices and ratios mentioned in Section 2.1.1 (see, e.g., Tridico (2018); Zhou et al. (2011); Liberati (2015); Celikay et al. (2016); Yang and Qiu (2016)). Some studies have used the lower and upper quintiles or deciles of income distribution as alternative measures (Saltz, 1995; Ata et al., 2019; Shao, 2021), but only a few have incorporated all quintiles or deciles of the income distribution (Škare and Stjepanovic, 2014; Saccone, 2021).

Notably, apart from the study by Kagalwala et al. (2021), which employed a log-ratio model to analyze inequality determinants in the U.S., there is limited research utilizing a compositional data approach to model income distribution. In their work, Kagalwala et al. (2021) used dynamic models of compositional dependent variables to investigate the impact of factors such as polarization, marginal tax rates, returns to labor and capital, and partisan control of Congress on income shares across groups from 1947 to 2014. This approach highlights the trade-offs and variations among income groups, providing a nuanced understanding of income inequality. As they noted, examining the composition of relative income shares offers a more comprehensive perspective on income inequality.

2.2 Compositional Data

Compositional data arise in almost all disciplines where researchers are interested in the dependence of non-negative proportions with unit-sum on certain relevant factors. The early recognition of CoDA has been demonstrated in the natural and health sciences such as geochemistry (Aitchison, 1984), biology (Campbell and Mosimann, 1987), archaeometry (Baxter and Freestone, 2006), and psychiatry (Gueorguieva et al., 2008). Later, more attention has been paid to the use of CoDA in other disciplines including safety (Tapiro et al., 2016, 2018), marketing (Sabnis et al., 2013; Morais et al., 2018), political science (Nguyen et al., 2022), public health and nutrition (Dumuid et al., 2018; Leite, 2019; Solans et al., 2019), economics (Brida et al., 2022), and education (Päuler-Kuppinger and Jucks, 2018).

2.2.1 Basics of Compositional Data

A composition is a positive vector $\mathbf{x} = (x_1, ..., x_D)$ whose components are subject to a constant sum constraint

$$x_1 + ... + x_D = \text{constant}.$$

A positive vector $\mathbf{w} = (w_1, ..., w_D)$ is compositional when our interest is in the relative magnitudes rather than the absolute values. The vector $\mathbf{w}$ is then transformed to a compositional vector $\mathbf{y}$ by using the closure operator $\mathcal{C}()$ as follows

$$\mathbf{y} = \mathcal{C}(\mathbf{w}) = \left(\frac{w_1}{t}, \frac{w_2}{t}, ..., \frac{w_D}{t}\right)$$

where $t = \sum_{i=1}^{D} w_i$.

The natural sample space of compositional data is the simplex $\mathcal{S}^D$;

$$\mathcal{S}^D = \{(x_1, \ldots, x_D) : x_j > 0 \text{ for } j = 1, \ldots, D \text{ and } \sum_{j=1}^{D} x_j = 1\}.$$

It is worth noting that, due to the constrained nature of the simplex, the traditional statistical techniques can not be used for analyzing compositional data (Aitchison, 1986).

2.2.2 Compositional Models

There are two common approaches in modeling compositional data analysis; the logratio model and Dirichlet regression. The logratio model, introduced by Aitchison (1986) in the 1980s, is based on transforming the compositional data from the constrained sample space $\mathcal{S}^D$ to the unconstrained sample space $\mathbb{R}^{D-1}$ using the additive logratio (alr) transformation. The resulting model is a multivariate linear model that facilitates statistical analysis within the unconstrained space. The alr transformation is defined as follows:

$$\mathbf{y} = (y_1, y_2, ..., y_D) \Longrightarrow \mathbf{w} = arl(\mathbf{y}) = \left(\ln\frac{y_1}{y_D}, \ln\frac{y_2}{y_D}, ..., \ln\frac{y_{D-1}}{y_D}\right) \tag{1}$$

where $\mathbf{y}$ represents the compositional data in the constrained sample space $\mathcal{S}^D$, and $\mathbf{w}$ represents the transformed data in the unconstrained sample space $\mathbb{R}^{D-1}$. In the transformed sample space, traditional statistical techniques, such as multivariate regression, can be reliably used to model the logratios using the explanatory variable(s) of interest (Egozcue and Pawlowsky-Glahn, 2019).

Unlike the logratio model, Dirichlet regression models the compositions based on the Dirichlet distribution in the simplex (Campbell and Mosimann, 1987; Hijazi and Jernigan, 2009). If $\mathbf{Y}$ =$(Y_1, ..., Y_D)$ has a Dirichlet distribution with positive parameters $(\alpha_1, ..., \alpha_D)$, then the density function of $\mathbf{Y}$ is given by

$$f(\mathbf{y}) = \left(\frac{\Gamma(\phi)}{\prod_{j=1}^{D} \Gamma(\alpha_j)} \right) \prod_{j=1}^{D} y_j^{\alpha_j - 1} = \left(\frac{\Gamma(\phi)}{\prod_{j=1}^{D} \Gamma(\phi\mu_j)} \right) \prod_{j=1}^{D} y_j^{\phi\mu_j - 1} \quad (2)$$

where $\sum_{j=1}^{D} y_j = 1$, $\phi = \sum_{j=1}^{D} \alpha_j$ is the dispersion parameter, and $\mu_j = \alpha_j/\phi$ is the mean of Y_j.

As shown in Equation (2), Dirichlet density could be presented in two forms. The first form is the *common parametrization* (Hijazi and Jernigan, 2009) while the second one is called the *alternative parametrization* (Maier, 2014). Under the common parametrization, each α_j could be written as a function of the explanatory variable(s). On the other hand, under the alternative parametrization, the expected values (μ_j's) and the precision parameter (ϕ) are modeled. In a GLM-like fashion, one component will be dropped and considered as the reference component. The required link functions would be defined as:

$$\begin{aligned} g_\mu(\mu) &= \mathbf{X}\beta \\ g_\phi(\phi) &= \mathbf{Z}\gamma \end{aligned}$$

The two Dirichlet regression models can be fitted using the '`DirichletReg`' package in R (Maier, 2021).

3 Models and Data

A Dirichlet regression model is used to identify the determinants of income inequality using cross-sectional country-level data. The dependent variable is the income distribution segmented into three categories (Poorest 20%, Middle 60%, Richest 20%) to capture the effects of explanatory variables on both the middle class and the tails (poorest and richest) of the income distribution.

The explanatory variables were selected based on their relevance to income inequality research, as discussed in Section 2.1.2. These variables include health disparities (adult mortality rate); education (primary school enrollment, government expenditures on education); labor dynamics (share of waged and salaried workers, unemployment rate); macroeconomic indicators (per capita GDP, GDP growth, inflation rate, government consumption expenditures); and institutional effectiveness indicators (control of corruption). See Table 1 for a detailed description of the model variables. Complete records of the variables were collected for 131 countries from the World Bank database for the most recent available year (World Bank, 2022a). Data for all the variables were obtained from the World Development Indicators database, except for the control of corruption which was extracted form the Worldwide Governance Indicators (World Bank, 2022b).

The log-link function given in Equation (3) is used to model the expected values of the lowest and the highest income quintiles while holding the share of the middle 60% of

Table 1: Variable abbreviation, description and source

Variable	Abbrev	Description
Share of the poorest 20%	QUIN1	The share of total income by the lowest 20% of population.
Share of the middle 60%	Mid60	The share of total income by the middle 60% of population.
Share of the richest 20%	QUIN5	The share of total income by the highest 20% of population.
Gini coefficient	GINI	Gini coefficient measured on a range from 0 to 100.
Adult mortality rate	MORT	Mortality rate per 1,000 adults.
Wage and salaried workers	WGE	Total wage and salaried workers expressed as a % of total employment.
Unemployment rate	UNEM	The % of workforce who are unemployed but looking for work.
Primary school enrollment	EDU	The primary school gross enrolment rate.
Government expenditure on education	GEDU	Expenditure on education as a % of GDP
Government consumption expenditure	GCON	Consumption expenditure as a % of GDP.
Per Capita GDP (log)	LGDP	The natural log of the GDP per capita.
Per Capita GDP Growth	GDPG	Annual % growth rate of GDP per capita.
Inflation rate	INF	Inflation as measured by the annual growth rate of the GDP.
Control of corruption*	CORR	Estimate of governance performance.

Source: World Development Indicators (WDI) and Worldwide Governance Indicators (WGI)*

the population as the reference component. The dispersion parameter (ϕ) is estimated as a constant.

$$\begin{aligned}\log(\mu_j(\mathbf{X})) = \quad & \beta_0 + \beta_1 MORT + \beta_2 WGE + \beta_3 UNEM + \beta_4 EDU + \beta_5 GEDU \\ & +\beta_6 GCON + \beta_7 LGDP + \beta_8 GDPG + \beta_9 INF + \beta_{10} CCOR\end{aligned} \tag{3}$$

The proposed Dirichlet regression model is compared with the OLS regression model (4), where the Gini coefficient is used as a dependent variable and the same explanatory variables, as outlined in Table 1, are included.

$$\begin{aligned}GINI = \quad & \beta_0 + \beta_1 MORT + \beta_2 WGE + \beta_3 UNEM + \beta_4 EDU + \beta_5 GEDU \\ & +\beta_6 GCON + \beta_7 LGDP + \beta_8 GDPG + \beta_9 INF + \beta_{10} CCOR + \epsilon\end{aligned} \tag{4}$$

4 Results and Discussion

Table 2 presents descriptive statistics for the income inequality measures and explanatory variables. The income share of the lowest quintile shows relatively limited variation compared to that of the highest quintile and the Gini coefficient. Specifically, the income share for the poorest 20% ranges from 2% to 11%, while the share for the richest

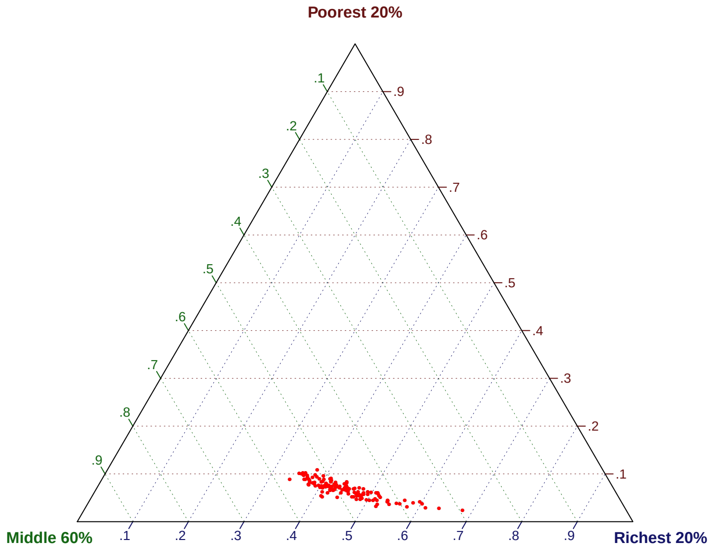

Figure 2: Ternary diagram of the income distribution of the selected countries

20% spans a broader range, from 34% to 68%. The control of corruption indicator, which typically varies between -2.5 (weak control) and 2.5 (strong control), falls within a narrower range in the selected countries, from -1.54 to 2.15. Among the explanatory variables, mortality rate and the percentage of wage and salaried workers exhibit the greatest variation across countries.

Figure 2 displays the distribution of income shares using a ternary diagram. This visualization shows that the shares of the poorest 20% exhibit minor variation across countries, whereas there is substantial variability in the shares of the richest 20%. At the extremes, Namibia and South Africa are positioned on the far right, illustrating the significant concentration of income within their wealthiest segments.

The results of the fitted models are presented in Table 3. The Dirichlet regression model is highly significant ($\chi^2 = 184.2$, $p < 0.001$), demonstrating a good fit to the income distribution, as evident in figures 3 and 4. All variables are significant at the 1% level, except for government expenditures on education, which is significant at the 10% level. The results highlight several significant explanatory variables with notable effects on income shares. GDP ($p = 0.047$), government consumption spending ($p = 0.017$), and control of corruption ($p = 0.077$) significantly impact the income share of the poorest 20%, though the control of corruption is significant only at the 10% level, indicating a weaker relationship. For the richest 20%, nearly all variables show highly significant effects. The “effect” column in the Dirichlet regression results provides marginal effects similar to coefficients in multiple regression models, offering valuable insights into the impact of explanatory variables on income shares. For instance, a one-point increase in control of corruption reduces the income share of the richest 20% by 2.05%, suggesting its

Table 2: Descriptive statistics of model variables (n=131)

Variable	Mean	Standard Deviation	Minimum	Maximum
QUIN1	0.07	0.02	0.02	0.11
Mid60	0.48	0.05	0.30	0.57
QUIN5	0.45	0.06	0.34	0.68
GINI	37.8	7.8	24.6	63.0
MORT	191.3	94.4	57.6	545.7
WGE	58.3	25.8	4.9	95.7
UNEM	7.37	5.65	0.50	28.47
EDU	104.4	11.9	66.4	143.7
GEDU	4.49	1.66	1.33	10.68
GCON	15.99	6.44	3.60	56.41
LGDP	8.63	1.43	5.57	11.65
GDPG	1.78	2.98	-9.40	16.42
INF	5.69	9.50	-2.85	61.31
CORR	-0.03	0.96	-1.54	2.15

potential role in mitigating inequality. Conversely, increased government consumption expenditure widens inequality, highlighting its disproportionate benefit to higher-income groups. These findings align closely with the OLS regression results shown in the last two columns of Table 3.

However, some variables exhibit limited or insignificant effects on the income share of the poorest 20%. For example, primary school enrollment ($p = 0.436$) and inflation rate ($p = 0.474$) show no meaningful impact, suggesting that broader educational reforms or inflation control may not directly influence income inequality without addressing structural issues. Similarly, the effects of unemployment rate ($p = 0.193$) and government expenditure on education ($p = 0.535$) are insignificant for the poorest 20%, though their marginal effects on the richest 20% remain notable. These findings suggest that while some indicators are critical for addressing income disparities, others require further investigation to understand their nuanced or indirect roles. Overall, the results provide valuable insights into the multifaceted drivers of income inequality and align with established literature in the field as discussed in Section 2.1.2.

5 Conclusion

In conclusion, the results of the Dirichlet regression and OLS models provide valuable insights into the dynamics of income distribution. Unlike the OLS regression, the Dirichlet regression model, with its compositional approach, offers clear interpretations of the effects of various indicators on the individual shares of income distribution. The findings demonstrate that government consumption, control of corruption, and GDP significantly

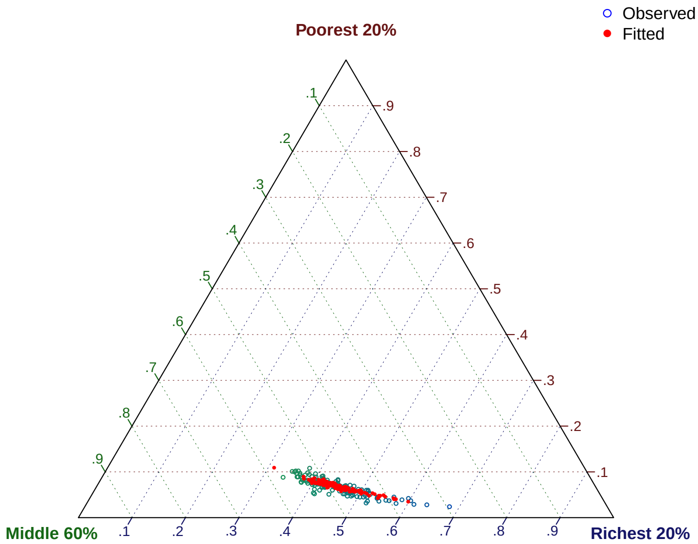

Figure 3: Ternary diagram for observed and fitted income distributions

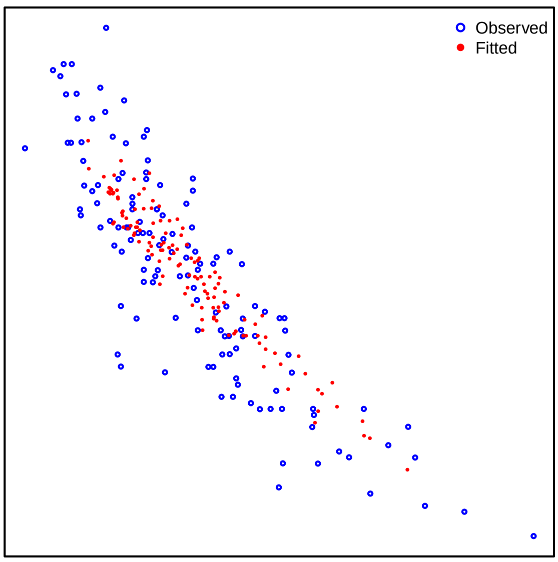

Figure 4: A zoom-out of the observed and fitted income distributions

Table 3: Results of Dirichlet regression and the OLS

	Dirichlet Regression						OLS Regression		
	Poorest 20%			Richest 20%					
Variable	Effect	p-value		Effect	p-value		Coefficient	p-value	
MORT	-0.01%	0.105		0.03%	<0.001	***	0.03%	<0.001	***
WGE	0.04%	0.062		-0.10%	0.002	***	-0.13%	0.003	***
UNEM	0.04%	0.193		-0.17%	<0.001	***	-0.17%	0.006	***
EDU	-0.02%	0.436		0.06%	0.062	*	0.07%	0.137	
GEDU[a]	-0.18%	0.535		0.83%	0.003	***	0.94%	0.031	**
GCON	-0.97%	0.017	**	2.07%	0.007	***	2.96%	0.004	***
LGDP	-0.10%	0.047	**	0.33%	<0.001	***	0.40%	<0.001	***
GDPG	0.15%	0.194		-0.59%	<0.001	***	-0.74%	<0.001	***
INF	0.05%	0.474		-0.21%	0.002	***	-0.23%	0.024	**
CORR	0.76%	0.077	*	-2.05%	0.005	***	-2.66%	0.010	***

Middle 60% is omitted (reference category)

Significance codes: '***' 0.01 '**' 0.05 '*' 0.1

[a] Significant at 10%. All other variables are significant at 1% level.

affect income inequality. For instance, corruption control reduces the income share of the wealthiest groups and thus mitigates inequality, while economic growth, as measured by GDP, benefits the poorest 20%. In contrast, government consumption spending exacerbates inequality, primarily benefiting the wealthiest groups. On the other hand, variables such as unemployment, inflation, and education exhibit nuanced or insignificant effects on the poorest income group, suggesting the need for further investigation into their indirect or structural impacts.

The proposed compositional approach to modeling income inequality, particularly through Dirichlet regression, provides a more detailed understanding of income shares compared to traditional measures like the Gini coefficient. This richer perspective supports the development of more effective policy recommendations. For example, policies targeting corruption control and fostering GDP growth could reduce inequality. Additionally, while government consumption spending currently exacerbates inequality, its strategic allocation could potentially benefit disadvantaged groups.

The accessibility of well-developed R packages such as `DirichletReg` enhances the usability of this approach, supporting model estimation and the visualization of compositional data. Complementary methods, such as the logratio model, could also provide further insights into income inequality.

Finally, while these models offer robust tools for understanding income distribution dynamics, more work is needed to develop advanced models for compositional panel data. Policymakers can leverage these findings to implement targeted interventions addressing specific drivers of inequality. Such evidence-based strategies could potentially promote inclusive economic growth, reduce income disparities, and ensure more equitable resource distribution.

References

Adams, S. and Klobodu, E. K. M. (2016). Financial development, control of corruption and income inequality. *International Review of Applied Economics*, 30(6):790–808.

Afonso, H., LaFleur, M., and Alarcón, D. (2015). Inequality measurement. *Development Issues*, 2.

Aitchison, J. (1984). The statistical analysis of geochemical compositions. *Journal of the International Association for Mathematical Geology*, 16:531–564.

Aitchison, J. (1986). *The statistical analysis of compositional data.* Chapman and Hall London.

Anderson, E., Jalles D'Orey, M. A., Duvendack, M., and Esposito, L. (2017). Does government spending affect income inequality? a meta-regression analysis. *Journal of Economic Surveys*, 31(4):961–987.

Asogwa, F. O., Onyeke, Q. O., Kuma, P. M., Arazue, W. O., and Nkalu, C. N. (2022). Do macroeconomic indicators determine income inequality in selected African countries? *Journal of Public Affairs*, 22(4):e2560.

Ata, F. N., Alam, S., and Saeed, N. (2019). The determinants of income distribution: an empirical analysis of developing countries. *Public Finance Quarterly*, 64(4):494.

Atkinson, A. B. (2015). Can we reduce income inequality in OECD countries? *Empirica*, 42(2):211–223.

Baxter, M. J. and Freestone, I. C. (2006). Log-ratio compositional data analysis in archaeometry. *Archaeometry*, 48(3):511–531.

Brida, J. G., Lanzilotta, B., and Moreno, L. (2022). Compositional tourists' expenditure: Modeling through Dirichlet regression. *Tourism Economics*, page 13548166221108819.

Campbell, G. and Mosimann, J. E. (1987). Multivariate analysis of size and shape: modelling with the Dirichlet distribution. In *ASA Proceedings of Section on Statistical Graphics*, pages 93–101.

Carvajal, C. R., Rodríguez, M. A., and Cuartas, B. M. (2019). Determinants of income inequality reduction in the latin american countries. *CEPAL Review No. 126, December 2018*, page 79.

Celikay, F., Celikay, F., Sengur, M., and Sengur, M. (2016). Education expenditures and income distribution: an empirical analysis on european countries. *Humanomics*, 32(3):248–257.

Ceriani, L. and Verme, P. (2012). The origins of the Gini index: extracts from variabilità e mutabilità (1912) by corrado Gini. *The Journal of Economic Inequality*, 10:421–443.

Choi, C. (2006). Does foreign direct investment affect domestic income inequality? *Applied Economics Letters*, 13(12):811–814.

Cobham, A. and Sumner, A. (2013). Is it all about the tails? the palma measure of income inequality. *Center for Global Development working paper*, (343).

Cysne, R. P. and Turchick, D. (2012). Equilibrium unemployment-inequality correlation. *Journal of Macroeconomics*, 34(2):454–469.

Davydov, Y. and Greselin, F. (2020). Comparisons between poorest and richest to measure inequality. *Sociological Methods & Research*, 49(2):526–561.

De Maio, F. G. (2007). Income inequality measures. *Journal of Epidemiology & Community Health*, 61(10):849–852.

Dorling, D. (2015). Income inequality in the UK: Comparisons with five large Western European countries and the USA. *Applied Geography*, 61:24–34.

Dumuid, D., Stanford, T. E., Martin-Fernández, J.-A., Pedišić, Ž., Maher, C. A., Lewis, L. K., Hron, K., Katzmarzyk, P. T., Chaput, J.-P., Fogelholm, M., et al. (2018). Compositional data analysis for physical activity, sedentary time and sleep research. *Statistical methods in medical research*, 27(12):3726–3738.

Egozcue, J. J. and Pawlowsky-Glahn, V. (2019). Compositional data: the sample space and its structure. *Test*, 28(3):599–638.

El Aboudi, S., Jouali, Y., El Bakkouchi, M., and Echaoui, A. (2024). Analyzing the dynamics of inflation, exchange rates and economic growth through the gini index: Modeling var in morocco. *International Journal of Economics and Financial Issues*, 14(6):136–144.

Erauskin, I. (2020). The labor share and income inequality: Some empirical evidence for the period 1990-2015. *Applied Economic Analysis*, 28(84):173–195.

Filiztekin, A. et al. (2015). Income inequality trends in Turkey. *Iktisat Isletme ve Finans*, 30(350):63–92.

Furceri, D. and Ostry, J. D. (2019). Robust determinants of income inequality. *Oxford Review of Economic Policy*, 35(3):490–517.

Ganaie, A. A., Bhat, S. A., and Kamaiah, B. (2018). Macro-determinants of income inequality: An empirical analysis in case of India. *Econ. Bull*, 38:309–325.

Gaston, N. and Rajaguru, G. (2009). The long-run determinants of australian income inequality. *Economic Record*, 85(270):260–275.

Gastwirth, J. L. (2017). Is the Gini index of inequality overly sensitive to changes in the middle of the income distribution? *Statistics and Public Policy*, 4(1):1–11.

Gil-Alana, L. A., Škare, M., and Pržiklas-Družeta, R. (2019). Measuring inequality persistence in oecd 1963–2008 using fractional integration and cointegration. *The Quarterly review of economics and finance*, 72:65–72.

Greselin, F., Pellegrino, S., and Vernizzi, A. (2021). The zenga equality curve: A new approach to measuring tax redistribution and progressivity. *Review of Income and Wealth*, 67(4):950–976.

Greselin, F. and Zitikis, R. (2018). From the classical gini index of income inequality to a new zenga-type relative measure of risk: A modeller's perspective. *Econometrics*, 6(1):4.

Gueorguieva, R., Rosenheck, R., and Zelterman, D. (2008). Dirichlet component regression and its applications to psychiatric data. *Computational statistics & data analysis*, 52(12):5344–5355.

Halkos, G. E. and Aslanidis, P.-S. C. (2023). Causes and measures of poverty, inequality, and social exclusion: A review. *Economies*, 11(4):110.

Haughton, J. and Khandker, S. R. (2009). *Handbook on poverty+ inequality.* World Bank Publications.

Hijazi, R. H. and Jernigan, R. W. (2009). Modelling compositional data using Dirichlet regression models. *Journal of Applied Probability & Statistics*, 4(1):77–91.

Hoffmann, F., Lee, D. S., and Lemieux, T. (2020). Growing income inequality in the United States and other advanced economies. *Journal of Economic Perspectives*, 34(4):52–78.

Hu, Y., van Lenthe, F. J., and Mackenbach, J. P. (2015). Income inequality, life expectancy and cause-specific mortality in 43 european countries, 1987–2008: a fixed effects study. *European Journal of Epidemiology*, 30(8):615 – 625.

Jedrzejczak, A. (2012). Estimation of concentration measures and their standard errors for income distributions in poland. *International Advances in Economic Research*, 18:287–297.

Kagalwala, A., Philips, A. Q., and Whitten, G. D. (2021). What about the rest of the pie? a dynamic compositional approach to modeling inequality. *Social Science Quarterly*, 102(4):1534–1552.

Kollmeyer, C. (2018). Trade union decline, deindustrialization, and rising income inequality in the United States, 1947 to 2015. *Research in Social Stratification and Mobility*, 57:1–10.

Kunawotor, M. E., Bokpin, G. A., and Barnor, C. (2020). Drivers of income inequality in africa: Does institutional quality matter? *African Development Review*, 32(4):718–729.

Leite, M. L. C. (2019). Compositional data analysis as an alternative paradigm for nutritional studies. *Clinical nutrition ESPEN*, 33:207–212.

Liao, T. F. (2022). Individual components of three inequality measures for analyzing shapes of inequality. *Sociological Methods & Research*, 51(3):1325–1356.

Liberati, P. (2015). The world distribution of income and its inequality, 1970–2009. *Review of Income and Wealth*, 61(2):248–273.

Liu, Y., Gastwirth, J. L., et al. (2020). On the capacity of the Gini index to represent income distributions. *METRON*, 78(1):61–69.

Maier, M. J. (2014). DirichletReg: Dirichlet regression for compositional data in R. Research report series / department of statistics and mathematics, WU Vienna University of Economics and Business, Vienna.

Maier, M. J. (2021). DirichletReg: Dirichlet regression in R. *R package version 0.7-1.*

Marrero, G. A. and Rodríguez, J. G. (2016). Macroeconomic determinants of inequality of opportunity in the united states: 1970–2009. In *Inequality after the 20th Century: Papers from the Sixth ECINEQ Meeting*, pages 307–330. Emerald Group Publishing Limited.

Morais, J., Thomas-Agnan, C., and Simioni, M. (2018). Using compositional and dirichlet models for market share regression. *Journal of Applied Statistics*, 45(9):1670–1689.

Nantob, N. et al. (2015). Income inequality and inflation in developing countries: An empirical investigation. *Economics Bulletin*, 35(4):2888–2902.

Ngamaba, K. H., Panagioti, M., and Armitage, C. J. (2018). Income inequality and subjective well-being: a systematic review and meta-analysis. *Quality of Life Research*, 27:577–596.

Nguyen, L., Kaizoji, T., and Inoue, T. (2024). The causal relationship between income inequality and economic growth: evidence from panel data for vietnam. *Applied Economics*, pages 1–15.

Nguyen, T. H. A., Laurent, T., Thomas-Agnan, C., and Ruiz-Gazen, A. (2022). Analyzing the impacts of socio-economic factors on french departmental elections with coda methods. *Journal of Applied Statistics*, 49(5):1235–1251.

Osberg, L. (2017). On the limitations of some current usages of the Gini index. *Review of Income and Wealth*, 63(3):574–584.

Päuler-Kuppinger, L. and Jucks, R. (2018). Out-of-class instruction in higher education–impact of approaches to teaching and discipline. *International Journal of Higher Education*, 7(2):199–209.

Petrakos, G., Rontos, K., Vavoura, C., and Vavouras, I. (2023). The impact of recent economic crises on income inequality and the risk of poverty in greece. *Economies*, 11(6):166.

Płatkowski, P. and Lechman, E. (2024). Reducing income inequalities through government expenditures on education: Evidence for european countries. *Applied Economics Letters*, pages 1–8.

Polacko, M. (2021). Causes and consequences of income inequality–an overview. *Statistics, Politics and Policy*, 12(2):341–357.

Porro, F. and Zenga, M. (2020). Decomposition by subpopulations of the zenga-84 inequality curve and the related index ζ : an application to 2014 bank of italy survey. *Statistical Methods and Applications*, 29(1):187 – 207.

Rebeira, M., Grootendorst, P., Coyte, P. C., and Aguirregabiria, V. (2017). Does rising income inequality affect mortality rates in advanced economies? *Economics*, 11.

Roser, M. and Cuaresma, J. C. (2016). Why is income inequality increasing in the developed world? *Review of Income and Wealth*, 62(1):1–27.

Ross, N. A., Dorling, D., Dunn, J. R., Henriksson, G., Glover, J., Lynch, J., and Weitoft, G. R. (2005). Metropolitan income inequality and working-age mortality: a cross-sectional analysis using comparable data from five countries. *Journal of Urban Health*, 82:101–110.

Sabnis, G., Chatterjee, S. C., Grewal, R., and Lilien, G. L. (2013). The sales lead black hole: On sales reps' follow-up of marketing leads. *Journal of marketing*, 77(1):52–67.

Saccone, D. (2021). Who gains from economic freedom? a panel analysis on decile income shares. *Applied Economics Letters*, 28(8):646–649.

Saltz, I. S. (1995). Income distribution in the third world: Its estimation via proxy data. *American Journal of Economics and Sociology*, 54(1):15–24.

Shahabadi, A., Nemati, M., and Hosseinidoust, S. E. (2018). The effect of education on income inequality in selected islamic countries. *International Journal of Asia-Pacific Studies*, 14(2).

Shao, L. F. (2021). Robust determinants of income distribution across and within countries. *PloS one*, 16(7):e0253291.

Sidek, N. Z. M. (2021). Do government expenditure reduce income inequality: evidence from developing and developed countries. *Studies in Economics and Finance.*

Signor, D., Kim, J., and Tebaldi, E. (2019). Persistence and determinants of income inequality: The Brazilian case. *Review of Development Economics*, 23(4):1748–1767.

Skare, M., Gil Alana, L. A., Claudio Quiroga, G., Pržiklas Družeta, R., et al. (2021). Income inequality in China 1952-2017: persistence and main determinants.

Škare, M. and Stjepanovic, S. (2014). Income distribution determinants and inequality–international comparison. *Amfiteatru Economic Journal*, 16(37):980–993.

Solans, M., Coenders, G., Marcos-Gragera, R., Castelló, A., Gràcia-Lavedan, E., Benavente, Y., Moreno, V., Pérez-Gómez, B., Amiano, P., Fernández-Villa, T., et al. (2019). Compositional analysis of dietary patterns. *Statistical methods in medical research*, 28(9):2834–2847.

Šoltés, E., Gawrycka, M., Szymczak, A., and Kuc-Czarnecka, M. (2023). Labour share and income inequalities in the european union, taking into account the level of development of economies. *Equilibrium. Quarterly Journal of Economics and Economic Policy*, 18(4):941–965.

Szymańska, A. and Zielenkiewicz, M. (2022). Declining labour income share and personal income inequality in advanced countries. *Sustainability*, 14(15):9403.

Tapiro, H., Oron-Gilad, T., and Parmet, Y. (2016). Cell phone conversations and child pedestrian's crossing behavior; a simulator study. *Safety Science*, 89:36–44.

Tapiro, H., Oron-Gilad, T., and Parmet, Y. (2018). The effect of environmental distractions on child pedestrian's crossing behavior. *Safety science*, 106:219–229.

Tridico, P. (2018). The determinants of income inequality in OECD countries. *Cambridge Journal of Economics*, 42(4):1009–1042.

World Bank (2022a). *World Development Indicators 2022.* World Bank Publications.

World Bank (2022b). *Worldwide Governance Indicators.* World Bank Data. Retrieved from https://info.worldbank.org/governance/wgi/.

Yang, J. and Qiu, M. (2016). The impact of education on income inequality and intergenerational mobility. *China Economic Review*, 37:110–125.

Zenga, M. (2007). Inequality curve and inequality index based on the ratios between lower and upper arithmetic means. *Statistica e Applicazioni*, 5(1):3 – 27.

Zhou, L., Biswas, B., Bowles, T., and Saunders, P. J. (2011). Impact of globalization on income distribution inequality in 60 countries. *Global Economy Journal*, 11(1).